3I2-OS-20-11

プライバシ保護クエリ方式の提案

A Proposal of Privacy-Preserving Query Scheme

伊藤 孝一[*1]
Kouichi ITOH

津田 宏[*2]
Hiroshi Tsuda

[*1] 株式会社 富士通研究所
セキュアコンピューティング研究部
Secure Computing Laboratory, Fujitsu Laboratories Ltd.

[*2] 株式会社 富士通研究所
ソフトウェアシステム研究部
Software System Laboratories, Fujitsu Laboratories Ltd.

ブラックリストを管理する業界クラウド等のデータベースに対し、信用調査のため自社の顧客を情報を問い合わせるシーンを考慮した場合、信用調査を行う会社は、自社顧客がブラックリスト該当者であることを業界クラウドを含めた外部に知られたくない。本稿は、自社顧客がブラックリスト該当者であることを外部に知られることなく、自社顧客のブラックリストに対する類似度を編集距離により算出可能な方式を提案する。

1. はじめに

クラウドコンピューティングを利用したデータ利活用が着目されている。異なる組織が持つデータを互いに利活用することで、単一組織のデータ利活用では得られなかった知見を得ることが期待される。一方、異なる組織をまたがったデータ利活用による情報の漏えいも懸念されており、プライバシ保護とデータ利活用を両立することが需要な課題である。

プライバシ保護とデータ利活用を両立するための技術は、プライバシ保護データマイニングと呼ばれている(以下 PPDM)。最初の PPDM 技術として、Rakesh らによって提案された手法が知られている[Rakesh00]。Rakesh らは、データに乱数を加算しマスク化することでプライバシ保護を実現しつつ、マスク化されたデータに対する分析手法を提案している。Rakesh らによって PPDM が提案されて以来、様々な PPDM 手法が提案されてきた[Ikarashi08, Ushida11 等]。これらの手法を用いることで、データは常に秘匿化された状態で各種の分析を行うことができるので、プライバシ保護とデータ利活用を両立することができる。

検索としての PPDM としては、プライバシ保護検索技術と呼ばれるものが知られており、暗号化したままの状態で検索を行うことで、異組織間で互いのデータを知ることなく検索を実現するプライバシ保護を実現する。しかし、暗号化したままの検索を行うために、様々な制約が発生する。例えば、完全一致しか判定できない[Katz08]、タグ付きのデータに対してしか検索できない[Boneh04, Matsuda12]]、という制約がある。さらに、検索対象の DB に対して暗号化を行う必要があり、既存の DB に対して修正を加える必要がある。これらの制約から、従来のプライバシ保護検索技術は、異組織間のデータ利活用に利用しにくいという側面があった。

1.1 本論文の貢献

本稿では、従来のプライバシ保護検索技術の問題を解決した、新しいプライバシ保護検索技術を提案する。本稿では、プライバシ保護検索技術の利用シーンとして、ブラックリストを管理する業界クラウド等のデータベースに対し、信用調査のため自社の顧客を情報を問い合わせるケースを想定する。このシーンにおいて、信用調査を行う会社は、自社顧客がブラックリスト該当者であることを業界クラウドを含めた外部に知られたくない。完全一致のみの検索であるならば、従来手法でも実現できるが、類似検索を必要とされる場合従来手法では実現できない。本稿は、編集距離を利用した類似検索に対応したプライバシ保護検索技術を提案する。本稿の手法を用いることで、業界クラウド側の DB を修正することなく、プライバシ保護検索を実現できる。

〒211-8588 神奈川県川崎市中原区上小田中 4-1-1
kito@labs.fujitsu.com, htsuda@jp.fujitsu.com

2. 利用シーンと従来法の問題点

2.1 利用シーン

本稿で想定する利用シーンを図1に示す。金融機関等が自社顧客の信用調査を行うために、債務ブラックリスト DB を保持している業界クラウドにクエリ送信を行う。業界クラウドは、ブラックリストを保持している複数の組織の DB から構成され、業界クラウドに対してクエリ送信を行うことで、これら複数の DB を横断した検索が行われ、編集距離による類似度比較が行わる。この結果、類似度が高い判定されたデータがクエリ送信元にレスポンスとして返信される。

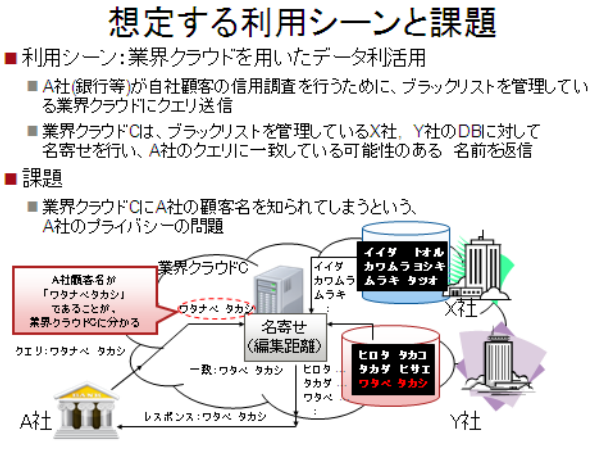

図1 想定する利用シーン

2.2 従来手法とその課題

想定される利用シーンでは、A 社が調べようとしている自社顧客名を業界クラウド C に対して秘匿することが、プライバシ保護を実現するための課題である。この課題に対しては、暗号を用いたプライバシ保護検索による解決策が考えられるが、完全一致の検索のみ可能であるため、類似検索ができないという問題がある。

第二の手法として、図 2 に示すように、クエリにダミー名を混在させる手法が考えられる。この手法を用いることで、類似検索にも対応できる。n 個のダミー名を用いた場合、業界クラウドは真の顧客名を含めた n+1 個の人名のうち、いずれが正しい名前であるかを識別できない。ただし、目的はクエリ元の顧客名がブラックリストに該当した場合、そのことを業界クラウドに知られないことである。従って、ダミーで追加した n 個の人名もブラックリスト該当者としてヒットすることが望ましい。しかし、このダミー人名をヒットさせるには、クエリ元が生成するダミーの顧客名を飛躍的に増大させる必要がある。

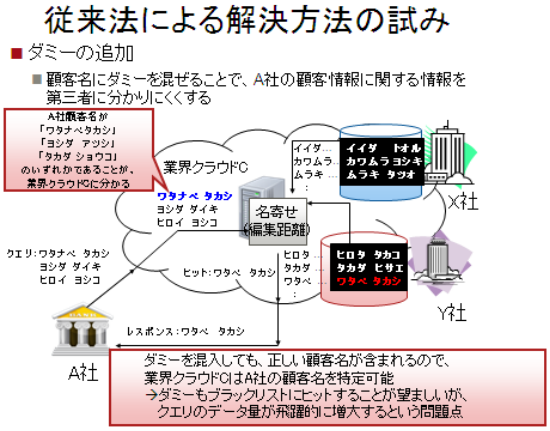

図2 ダミー名の追加による解決法とその問題点

3. 提案手法

提案手法の概要を図 3 に示す。クエリを送信する際に、ダミー名を用いる点は従来法と同じであるが、提案手法では、調査対象の顧客名とダミー名を文字単位で混合させる点が異なる。

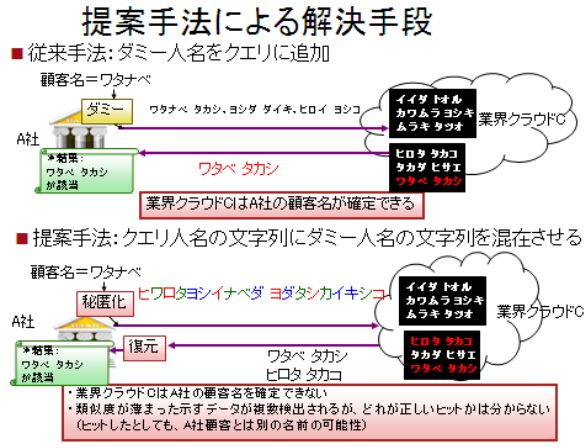

図 3 提案手法の概要

提案手法を用いた場合、業界クラウドでは検索ヒット条件となる類似度の閾値を下げる必要がある。クエリ送信元は、業界クラウドから複数のブラックリスト該当者データを受け取るが、これらのうち調査対象の顧客名と類似するものを選別する。

3.1 従来手法と提案手法の比較

n 個のダミー名を用いた場合の業界クラウドから見たクエリ元の人名の候補数は、従来手法の場合 n+1 個であった。これに対し提案手法では、調査対象の顧客名を n 個の人名を混合させる。この混合により、カナ表記で姓、名をクエリ送信する場合の候補数は、以下のように表される。

$$\Sigma_{i=i\_min,\ldots,i\_max}\left(p(i)\ {}_{ns}C_i\right) \times \Sigma_{j=j\_min,\ldots,j\_max}\left(q(j)\ {}_{ng}C_j\right) \quad (1)$$

ただし、s は姓の平均カナ文字数、g は名の平均カナ文字数である。i_min は姓のカナ文字数の最小数、i_max は姓のカナ文字数の最大数、p(i)は i 文字のランダムな文字が姓に一致する確率である。また、j_min は名のカナ文字数の最小数、i_max は名のカナ文字数の最大数、q(j)は j 文字のランダムな文字が名に一致する確率である。

B 人に一人がブラックリスト該当者である場合、ダミー名がブラックリストに該当する人数は従来法では n/B であるのに対し、提案法では、式(1)を B で割った人数、すなわち

$$\left(\Sigma_{i=i\_min,\ldots,i\_max}\left(p(i)\ {}_{ns}C_i\right) \times \Sigma_{j=j\_min,\ldots,j\_max}\left(q(j)\ {}_{ng}C_j\right)\right)/B$$

となるので、ダミー名のブラックリストへのヒット率を飛躍的に向上させることができる。

また、文字単位での共起率を考慮したダミー名混在を行うことで、確率 p(i),q(j)を高め、プライバシ保護を強化できると考える。

4. まとめ

業界クラウド等のデータ利活用に適した、プライバシ保護クエリ方式を提案した。提案方式は、編集距離を用いた類似検索に用いることができる。

参考文献

[Agrawal00] R.Agrawal and R.Srikant, “Privacy-Preserving Data Mining”, ACM SIGMOD Conference 2000, pp.439-450, 2000.

[Boneh04] D.Boneh, G.D.Cresecenzo, R.Ostrovsky, and G.Persiano, “Public key encryption with keyword serach”, Eurocrypt 2004, LNCS 3027, pp.506-522, 2004

[Golle04] P.Golle, J.Staddon, and B.Waters, “Secure conjunctive keyword search over encrypted data”, ACNS 2004, 2004.

[Ikarashi08] 五十嵐,千田,高橋, “多値属性に適用可能なプライバシ保護クロス集計”, CSS 2008, 2008.

[Katz08] J.Katz, A.Sahai, and B.Waters, “Predicate encryption supporting disjunctions, polynomial equations, and inner products”, Eurocrypt2008, LNCS 4965, pp.146-162, 2008

[Matsuda12] 松田,服部,平野,森,伊藤,川合坂井,太田, “安全性と高速性の両立を目指した検索可能暗号(1)”, SCIS 2012, 1A3-1, 2012.

[Ushida11] 牛田,伊藤,小櫻,津田, “ゲートウェイによるクラウド間のデータ秘匿集計技術”, SCIS 2011, 3F1-5, 2011